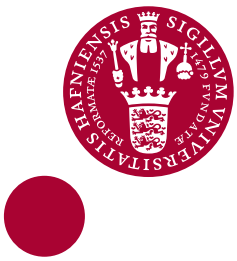


Semi-automatic segmentation of vertebrae in lateral X-rays using a conditional shape model

Iglesias, Juan Eugenio; de Bruijne, Marleen

Published in:
Academic Radiology

Publication date:
2007

Document version
Publisher's PDF, also known as Version of record

Semiautomatic Segmentation of Vertebrae in Lateral X-rays Using a Conditional Shape Model[1]

J. Eugenio Iglesias, MSc, Marleen de Bruijne, PhD

Rationale and Objectives. Manual annotation of the full contour of the vertebrae in lateral x-rays for subsequent morphometry is time-consuming. The standard six-point morphometry is commonly used instead. It has been shown that the information from the complete contour improves the quality of the study. In this article, the six landmarks are given and the vertebrae are segmented taking advantage of that information. The result is a semiautomatic system in which the full contour is found with high precision, and that only requires a radiologist to mark six points per vertebra.

Materials and Methods. A shape model was built for both the six landmarks and the full contours of the vertebrae L1, L2, L3, and L4 of 142 patients. The distribution of the principal components of the full contour was then modeled as a Gaussian conditional distribution depending on the principal components of the six landmarks. The conditional mean was used as initialization for active shape model optimization, and the conditional variance was used to constrain the solution to plausible shapes.

Results. The achieved point-to-line error was 0.48 mm, and 95% of the points were located within 1.36 mm of the annotated contour. The accuracy is superior to those of previously published studies, at the expense of requiring the six points to be marked. Fractures and osteophytes are well approximated by the model, although they are sometimes oversmoothed.

Conclusions. The proposed method provides hence a richer description than the six points, and can be used as input for shape-based morphometry to detect and quantify vertebral deformation more accurately.

Key Words. Osteoporosis; lateral x-rays; vertebral fracture; vertebra segmentation; shape model.

Osteoporosis is a disease of bone in which the bone mineral density is reduced, bone microarchitecture is disrupted, and the amount and variety of noncollagenous proteins in bone is altered (1). Bones affected by the disease are more likely to fracture. Osteoporosis is defined by the World Health Organization as either a bone mineral density 2.5 standard deviations below peak bone mass (20-year-old, sex-matched healthy person average) as measured by dual x-ray absorptiometry (DXA), or any fragility fracture. Because of its hormonal component, more women, particularly after menopause, suffer from this disease than men.

Osteoporotic fractures are those that occur under slight amount of stress that would not normally lead to fractures in nonosteoporotic people. Typical fractures occur in the vertebral column, hip, and wrist. Vertebral fractures are the most common ones. They occur in younger patients and they are a good indicator for the risk of future spine and hip fractures. These two are the most serious cases, leading to limited mobility and possibly disability. Hip fracture, in particular, usually requires major surgery, which has important associated risks, such as deep vein thrombosis and pulmonary embolism. Although osteoporosis patients have an increased mortality rate because of the complications of fractures, most patients die with the disease rather than of it.

Acad Radiol 2007; 14:1156–1165

[1] From the Department of Computer Science, University of Copenhagen, Denmark (J.E.I., M.d.B.), and Nordic Bioscience, Herlev, Denmark (M.d.B.). Received January 9, 2007; accepted June 3, 2007. **Address correspondence to:** J.E.I. e-mail: marleen@diku.dk

doi:10.1016/j.acra.2007.06.003

Vertebral fractures are conventionally detected and graded on lateral x-rays. Apart from the subjective judgment of the image by a radiologist, the standard six-point morphometry (2,3) is commonly used. In this technique, six landmarks are placed on the corners and in the middle point of both vertebra endplates, defining the anterior, middle, and posterior heights. These measurements can be used to calculate a fracture grade, as in previous work (4,5). In current clinical trials, a fractured vertebra is defined as the one for which one of the three heights is at least 20% larger than any other.

The six-point representation captures most of the important information in the image, but it is unable to describe certain structures (such as osteophytes) or subtle shape variations. In earlier work (6), a point distribution model (PDM) was used to represent the full contour of vertebrae, and a classifier to separate normal and fractured vertebrae. An improvement was observed in the performance of the full contour representation compared to that of the six landmarks. Full-contour based morphometry was also carried out in (9). Because annotating the full contour of the vertebrae represents a huge load of work, many methods have been proposed to segment them automatically.

Automatic segmentation of the vertebrae in lateral x-ray images is a difficult task because of the nature of such images: they show the superimposition of many layers, making it difficult to distinguish the region of interest. This region is just a sagittal plane along the spine. This is the reason why many classic segmentation approaches, such as those based on region growing or snakes, fail. Active shape models have the advantage that they make use of prior knowledge on vertebral shape and appearance and therefore do not need to rely completely on the information in the images.

Many fully and semiautomated, shape-model based segmentation attempts are described in the literature. In a fully automated system, Zamora et al (7) used an active shape model (ASM) including gray-level edge information, initializing it with a customized implementation of the generalized Hough transform. They applied the method on x-ray images and achieved errors lower than 6.4 mm in 50% of the lumbar images. Smyth et al (6) used ASM methods in a semiautomated system to segment vertebrae in DXA. Requiring the user to mark three points corresponding to the midpoints of the bottom of L4, top of T12, and top of T7, they achieved a root mean square (RMS) point-to-line error lower than 1.23 mm in 95% of the cases in healthy vertebrae, and lower than 2.24 mm in 92% of the fractures. De Bruijne et al (8) proposed a fully automated method based on shape particle filtering, lowering the average point-to-line error to 1.4 mm in radiographs.

Roberts et al (10) incorporated an active appearance model and a dynamic ordering algorithm to segment the vertebrae in DXA images. Requiring the user to annotate the same three points as Smyth et al, they achieved a point-to-line error of 0.79 mm. The best results in the current literature have been achieved by the same authors (11), who recently reported a 0.64-mm mean point-to-line error (95% of the points 2 mm within the contour) in healthy vertebrae on radiographs, and 1.06 mm (with 87% within 2 mm) in fractures. This time they required the clinician to mark the approximate center of each vertebra. They improved also their algorithm in DXA images (12) by using two different initializations, one assuming a severe fracture and other assuming a normal vertebra, and then picking the best solution. They required once more the centers of the vertebrae to be marked, and achieved 0.69 mm mean point-to-line errors in normal vertebrae and 0.96 mm in fractures.

The cited studies require little or no user intervention, which is obviously an advantage. Their main shortcoming is that, even if the ASM constraints ensure plausible resulting shapes, the uncertainty in the pose parameters makes it difficult to fit the model. In this article, it is proposed to make use of the information from the six landmarks (of the six-point morphometry, annotated by a radiologist), leading to a semiautomatic algorithm. At the expense of requiring slightly larger user intervention, the segmentation accuracy can be improved.

In this study, two PDMs were built from the manually annotated training images: one for the six landmarks and another for the full contour, which is represented by 67 points. It is then possible to study the relationship between them and build a conditional PDM for the full contour depending on the six points. One can then work with the conditional distribution, taking its mean for the initial solution and using its covariance, which will be "smaller" than that of the general model, to set the plausibility limits for the ASM.

For example, if the general model was used, a shape close to the mean but with the points corresponding to the six landmarks far away from the annotations would have a high likelihood. In the conditional model, moving the points corresponding to the six landmarks away from the six annotations will make the likelihood decrease very quickly. Using the information from the six landmarks helps improve

the accuracy of the segmentation and makes the probability of achieving a very bad result in a difficult case very small.

MATERIALS AND METHODS

Available Data

Full contours and sets of six landmarks from 142 patients were available. In case of the vertebrae having "double contours" from projection effects, the lower one was always chosen. Vertebrae L1, L2, L3, and L4 were analyzed, so 568 vertebrae (64 fractures) were included in the study. The images were 12-bit deep and their resolution was 570 dpi. They were stored in DICOM format. Because the application did not require such high resolution, they were downsampled by a factor of five after applying an anti-aliasing Gaussian filter. In a digital image, the maximum digital frequency that can be represented is equal to π. Because this frequency is going to be divided by five because of the downsampling, the discrete cutoff frequency of the anti-aliasing filter, defined as twice the standard deviation of the Gaussian, was made equal to $\omega_c = \pi/5$. This is equivalent to a standard deviation of $\sigma = 3.24$ pixels in the spatial domain.

The six landmarks and the contours were marked by three different radiologists. The radiologist who marked the landmarks on an image would always annotate the contour, too. In the annotation of the six landmarks, the corners were marked first. Then the perpendicular bisector of the segment joining the upper corners was displayed. It serves as a guide for the radiologist, who is supposed to place a landmark on the point of minimum height, and if it is unclear, as close to the bisector as possible. The process is then repeated for the lower plate. The displayed bisectors help the radiologists be consistent throughout the annotation process, minimizing the impact of interobserver variability in the PDM.

To annotate the whole contour, the radiologists drew a polygonal line with as many vertices as they wanted. This contour was the ground truth for the study. The six landmarks and the contours were annotated in different passes without showing the earlier annotation, so they do not necessarily overlap (Fig 1).

METHODS

The proposed method is based on the following steps:

1. Training:
   a. An expert radiologist annotates the six landmarks and the contours on the training images.
   b. Align the vertebrae in the dataset.
   c. A PDM is created for both the six landmarks and the full contour.
2. Segmentation
   a. A radiologist marks the six landmarks of the actual image to be segmented.
   b. The conditional model is built based on the position of the six landmarks on the image to analyze and on the two models derived from the training set.
   c. An initial solution is estimated, using the mean of the conditional model.
   d. Active shape modeling is used in order to fit the contour to the vertebra in the image, using the conditional covariance to constrain the solution.

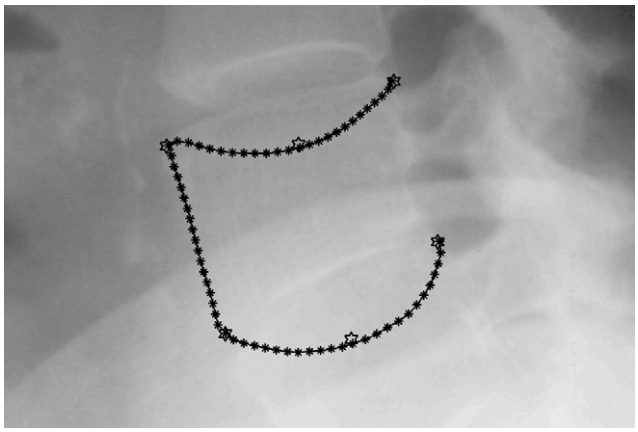


Figure 1. Six initial landmarks (stars), contour with and selected points (asterisks). Note that the six landmarks are not exactly on the contour.

Landmark Placement

Because a point distribution model is to be built, a set of landmarks must be defined for each available complete contour. As opposed to the six landmarks, the contours consist of a variable number of points, so it is necessary to resample them to ensure that the number of points is the same for every case. Because the maximum number of points annotated by a radiologist was 53, it was arbitrarily decided that the model would consist of 67 landmarks. The points of the contour closest to the initial six landmarks were chosen to be points 1, 13, 25, 43, 55, and 67. The rest of the landmarks were equidistantly placed between these six. The third segment has 50% more landmarks because it is on average (approximately) 50% longer than the other four. A sample image is shown in Fig 1. No resampling is required for the six points.

Alignment

Translation, scale, and rotation deviations between point sets must be eliminated through alignment. The Procrustes method (13) was used, using only the information from the four corners points from the six landmarks. Translation, rotation, and scaling, represented by four parameters, were allowed. The transform parameters for each vertebra were then applied on the corresponding full contours, so that both representations were aligned on the basis of the vertebral corners. Each aligned shape is then represented by a vector of landmark coordinates:

$$\mathbf{x} = [x_1,\ x_2,\ \ldots, x_N,\ y_1,\ y_2,\ \ldots, y_N]^t \qquad (1)$$

where $N = 6$ or $N = 67$, depending on the model.

PDM

After the training shapes are aligned, a point distribution model (14) is built for both the six landmarks and the full contour representations, mixing vertebrae L1, L2, L3, and L4 into a single model. To build the model, a principal component analysis was performed on the aligned data vectors. Principal component analysis is a technique, which can be used to reduce the dimensionality of a dataset. It is a linear transformation toward a new coordinate system such that the greatest variance by any projection of the data lies on the first coordinate (known as the first principal component), the second greatest variance on the second coordinate, and so on. It is therefore possible to simplify the dataset by keeping only the first few principal components in the new representation. The principal components, which constitute an orthonormal base, are defined by the eigenvectors of the covariance matrix of the dataset.

The shape model is then characterized by a mean shape $\bar{x}$, the kept eigenvalues λ_i, $i = 1, \ldots, k$, and the corresponding eigenvectors, which are grouped columnwise in the $\mathbf{P}$ matrix, which accomplishes that $\mathbf{P}^t\mathbf{P} = I$. Each shape can then be approximated by

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \qquad (2)$$

where b can be calculated for a certain shape:

$$\mathbf{b} = \mathbf{P}^t(\mathbf{x} - \bar{\mathbf{x}})$$

and the approximation error vector is:

$$\mathbf{e} = \mathbf{x} - \bar{\mathbf{x}} - \mathbf{P}\mathbf{b} = \mathbf{x} - \bar{\mathbf{x}} - \mathbf{P}\left(\mathbf{P}^t(\mathbf{x} - \bar{\mathbf{x}})\right)$$

where $\mathbf{b}$ is a column vector of k components, representing the projection of the shape onto the space of the model. Across the training set, the mean of this vector will be zero, and the covariance $\mathbf{C}$ will be a diagonal matrix including the k eigenvalues λ_i.

To verify whether a certain vector $\mathbf{b}$ corresponds to a plausible shape, it must be checked that it is not too far away from the mean of the model, that is, the zero vector. At the same time, plausible shapes can be generated just by taking $\mathbf{b}$ vectors close to the zero vector. The valid region is defined by limiting the Mahalanobis distance of $\mathbf{b}$. The limit d_{max} can for instance be chosen using the χ^2 distribution.

$$d = \sqrt{\mathbf{b}^t\mathbf{C}^{-1}\mathbf{b}} = \sqrt{\sum_{i=1}^{k}\left(\frac{b_i^2}{\lambda_i}\right)} \qquad (3)$$

If the condition $d < d_{max}$ is not true, the b vector can be modified:

$$\mathbf{b}' = \mathbf{b}(d_{max}/d) \qquad (4)$$

Because the number of fractures in the dataset is low compared with the number of healthy ones, their influence on the model was increased by giving them a higher weight when building it. Two different weights were given to normal and fractured vertebrae when calculating the mean and the variance of the shapes, so that their total contributions were equal. Because 504 healthy and 64 fractured vertebrae were available, the weights were

$$\mathrm{W_h} = \frac{1}{2N_{heal}} = \frac{0.5}{504}$$

$$\mathrm{W_f} = \frac{1}{2N_{frac}} = \frac{0.5}{64}$$

$$\bar{\mathbf{x}}' = \sum_{i=1}^{N_{heal}} w_h \mathbf{x}_i + \sum_{j=1}^{N_{frac}} w_f \mathbf{x}_j$$

$$\mathbf{C}' = \frac{N_{heal} + N_{frac}}{N_{heal} + N_{frac} - 1}\left\{\sum_{i=1}^{N_{heal}} w_h(\mathbf{x}_i - \bar{\mathbf{x}}')(\mathbf{x}_i - \bar{\mathbf{x}}')^t + \sum_{j=1}^{N_{frac}} w_f(\mathbf{x}_j - \bar{\mathbf{x}}')(\mathbf{x}_j - \bar{\mathbf{x}}')^t\right.$$

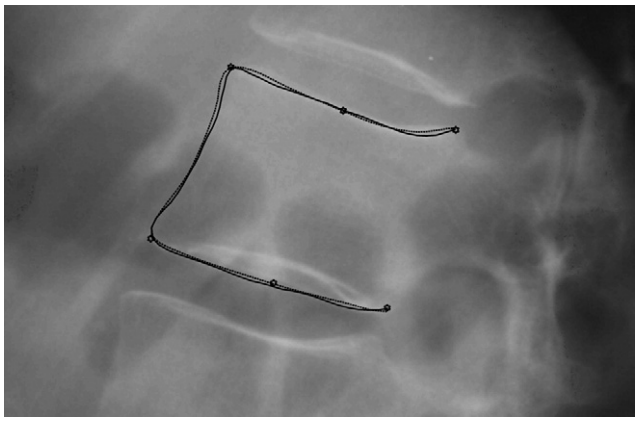


Figure 2. Initialization (dotted line) and real solution (solid line). The six landmarks are also marked.

Conditional Model

The distribution of the principal components of the full-contour model F can be modeled as a conditional Gaussian, dependent on the principal components of the landmarks L. If μ_F, μ_L, Σ_F, Σ_L are the means and covariances of the principal components for the two models across the training data, and Σ_{FL}, Σ_{LF} represent their cross-covariances, it is possible to write:

$$P(F \mid L) = N(b_{cond},\ C_{cond})$$
$$b_{cond} = \mu_F + \sum_{FL} \sum_{LL}^{-1} (L - \mu_L)$$
$$C_{cond} = \sum_{FF} - \sum_{FL} \sum_{LL}^{-1} \sum_{LF}$$

where b_{cond} is the conditional mean and C_{cond} the conditional covariance matrix for the principal component coordinates of the full contour given the principal component coordinates L of the six points.

It is also possible to model the position of the points and landmarks instead of using the principal components. In this case, multicollinearity in the positions might make Σ_{LL} noninvertible, and regularization can be required. This approach would lead to a more continuous downweighting of less important variation rather than using a cutoff at a fixed number of modes, but has the disadvantage that the sizes of the Σ_{LL}, Σ_{LF}, and Σ_{FL} matrices become larger.

The mean of the conditional model b_{cond} can be used as initialization (Fig 2), and the covariance will be very useful when fitting the model to the images. The conditional covariance C_{cond} is in general much "smaller" than the unconditional covariance C'. The differential entropy of the distribution decreases almost 10 logarithmic units from the unconditional to the conditional model (from –13.88 to –23.85). It is thus possible to look for the solution around the conditional mean, in a region limited by a certain value of Mahalanobis distance defined according to the new conditional covariance. The search space will hence be reduced, making it easier to fit the model and making shapes relatively far away from the six landmarks unlikely.

In the training data, that is, the annotated images, the six landmarks are not constrained to stay on their corresponding points on the contour. The lack of this constraint allows the 11-D full contour conditional shape model to represent exactly the same shapes as the nonconditional one, although with higher Mahalanobis distances and therefore smaller likelihood. This is important because it allows the conditional covariance matrix to remain full-rank, and thus invertible.

ASM

ASM (14) is an iterative algorithm that tries to fit the shape model to the contours of the vertebrae in the image. The first step is to find the translation (t_x, t_y), rotation (θ) and scale (s) parameters that best fit the corners of the six given landmarks to the corners of the mean of the shape model. These parameters define the transform that allows to switch between the positions of the points in the image X (in "physical" coordinates), and their positions in the shape model "normalized" coordinates x. These pose parameters will be kept constant throughout the process.

Starting from an initial solution (the conditional mean b_{cond}), which is calculated in the "normalized" coordinates with the six landmarks, a translation along the normal to the contour is proposed for every point in the model at each iteration. To calculate the optimal displacement, a gray value profile is built by sampling the gray levels in the image along the normal to the contour across each point. Instead of using these values directly, their normalized derivatives are calculated for the profiles: the increments are scaled so that the sum of their absolute values for each profile is one. The values are then scaled (for each profile) so that the sum. This makes the algorithm robust against contrast variations.

The resulting profiles are then compared to the ones of the training cases on the contour at the same landmark. If $p_i(t)$ represents the vector of normalized derivatives around point t in the profile around landmark i (let us say in the interval $[T - T_t,\ t + T_t]$), a fitness function $f_i(t)$ (15) can be calculated for each t by comparing $p(t)$ with the model built from the training examples (with semilength T_t):

$$f_i(t) = (\mathbf{p}_i(t) - \bar{\mathbf{p}}_i)^t S_{p,i}^{-1}(\mathbf{p}_i(t) - \bar{\mathbf{p}}_i)$$

where $\bar{\mathbf{p}}_i$ is the average of the profiles of length $2T_t + 1$ around the points in the training cases, and $S_{p,i}$ is a diagonal matrix including the (independent) variances of each element in the profile. The function $f_i(t)$ will just be defined in the interval $[-T_p, T_p]$. Therefore T_p defines how far from the current point location the search for the contour is performed. If t_m minimizes $f_i(t)$, the shift $0.66t_m$ is chosen to make the algorithm evolve in a smoother way.

If the displacements that minimize different $f_i(t)$ are stored in a vector dX by stacking the x and y coordinates in the same way as then shapes in Eq 1, it is possible to represent the new desired shape as $X + dX$. This suggested shape is then transformed into the normalized coordinates, becoming $x + dX$. The shape model parameters b are then updated to fit $x + dX$ as well as possible:

$$\left(\mathbf{P}^t\mathbf{W}_s\right)\mathbf{dx} = \left(\mathbf{P}^t\mathbf{W}_s\mathbf{P}\right)\mathbf{db}$$

where W_S is a diagonal matrix with weights that measure the importance of each point in the fitting. The weights depend on the magnitude of the displacement (15) and on the goodness of the fit:

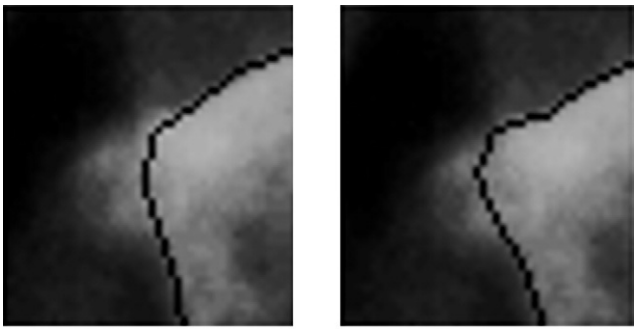
$$w'_i = \frac{1}{2 + \left|dX_i\right|^2}\frac{1}{f_i(t_m)}$$

$$w_i = \frac{w'_i}{\sum_{l=1}^{67} w'_i}$$

The idea behind the first term of the weight $(2 + |dX_i|^2)^{-1}$ is to prevent the shape model from fitting distant points that might be wrong, as least-squares is not robust against outliers. The second term $(f_i(t_m))^{-1}$ gives higher priority to fitting those points that are probably located on the right contour.

Before updating the contour, it is important to check that **db** leads to a plausible shape. It is now that the information in the covariance of the conditional model becomes useful. The main differences with Eq 3 and 4 are that the Mahalanobis distance is measured from the conditional mean $\mathbf{b}_{cond}$ instead of the origin and that the conditional covariance $\mathbf{C}_{cond}$ is not diagonal, so

$$\mathbf{b}^t\mathbf{C}^{-1}\mathbf{b} = \sum_{i=1}^{k}\left(\frac{b_i^2}{\lambda_i}\right)$$



a. **b.**

Figure 3. Influence of maximum allowed Mahalanobis distance on the result. In **(a)** the shape model is unable to fit the contour to the osteophyte. In **(b)**, the threshold has been increased by 1.5 and the contour approximates the osteophyte better. The shape model tends to oversmooth the osteophytes.

does not hold any longer (even though it would be possible to diagonalize $\boldsymbol{C}_{cond}$ and then use this simplification).

$$\mathbf{b}'_{i+1} = \mathbf{b}_i + \mathbf{db}$$

$$d = \sqrt{(\mathbf{b}'_{i+1} - \mathbf{b}_{cond})^t\mathbf{C}_{cond}^{-1}(\mathbf{b}'_{i+1} - \mathbf{b}_{cond})}$$

and then

$$\mathrm{b}_{i+1} = \begin{cases} \mathbf{b}'_{i+1}, & \text{if } d \le d_{\max} \\ \mathrm{b}_{cond} + (\mathbf{b}'_{i+1} - \mathbf{b}_{cond})(d_{\max}/d), & \text{if } d > d_{\max} \end{cases}$$

Therefore d_{max} is the parameter that controls how free the algorithm is to fit the contour to the edges in the image. A large value allows the result to move around the principal component space, which can lead to implausible solutions if the edges are not clear in the image. A small value makes the algorithm rely mostly on the model, leading to more conservative solutions, closer to the mean of the distribution. This can prevent the algorithm from finding the correct solution, especially in abnormal cases with fractures or osteophytes, in which the real contour is relatively far from the initialization in the principal component space (Fig 3).

The new coordinates are easily calculated using Eq 2 and then transformed back to the physical coordinates by the transform defined by $-t_x$, $-t_y$, $-\theta$, s^{-1}. A new translation is then proposed for each point once more, starting a new iteration. When the difference between the shapes at two consecutive iterations is smaller than a certain limit ($\|\mathrm{dX}\| \le 1$ pixel), the process stops.

RESULTS

Parameter Setting

First of all, several preliminary experiments were run to find suitable values for the different parameters in the algorithm. Regarding the selection of the number of principal components to be kept, it depends on the proportion of the variance to be preserved. Seven and eleven components, for the six landmarks and the full contour, respectively, were enough to keep approximately 97% of the total variance. From that point on, raising that proportion becomes very expensive in terms of the number of principal components.

Regarding the profile lengths, it was found that using a gray value model for the profile with semilength $T_t = 12$ pixels (2.7 mm) and making $T_p = 8$ pixels (1.8 mm) leads to good results. T_p represents how far from the current solution one tries to find the contour. Making this parameter too large would make the search region too large and hence make it more likely that the algorithm captures a wrong edge, especially if this edge does not represent an implausible shape. This happens typically in "double contour" cases (Fig 5b).

Finally, the choice of the maximum Mahalanobis distance d_{max} was made with the help of the χ^2 distribution for eleven degrees of freedom. Arbitrarily fixing a 10% tail probability, a compromise limit equal to 4.1 was picked. All these parameters were kept fixed for all further experiments, unless mentioned otherwise.

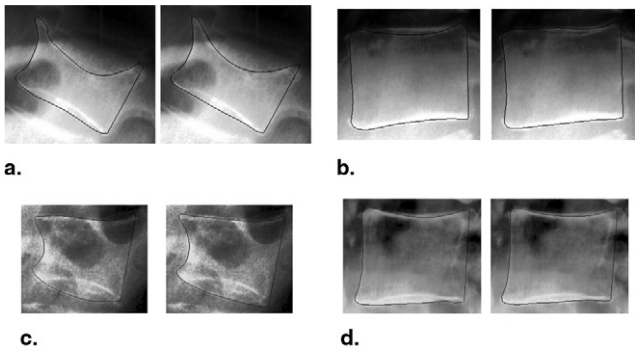


Figure 5. Some samples from the leave-one-out experiments. For each pair, the image on the left represents the ground truth, while the one on the right represents the output of the algorithm. **(a)** Oversmoothing in fractured vertebra (0.85 mm mean point-to-line error), **(b)** wrong edge captured (0.83 mm), **(c)** well-segmented fracture (0.51 mm), and **(d)** typical, normal, well-segmented vertebra (0.33 mm).

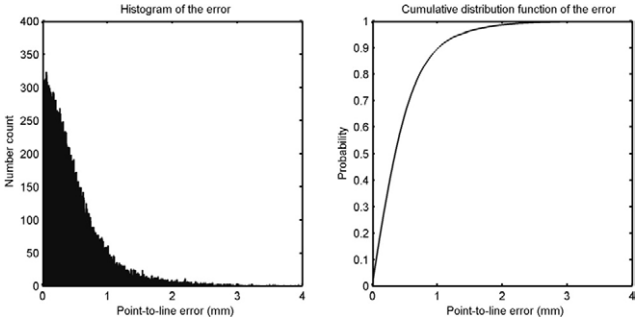


Figure 4. Distance to real contour: histogram and cumulative distribution function.

Evaluation

A leave-one-out experiment was carried out. For each vertebra, a complete model (six landmarks, full contour, profiles) was built from the other images and used for the segmentation. The distance between each point in the physician-annotated contour and the closest point in the detected contour (point-to-line distance) was calculated for each vertebra. The distribution of the error is represented in Fig 4.

The RMS error was equal to 0.68 mm, whereas the mean error was 0.48 mm and its standard deviation was 0.48 mm. A total of 89% of the points was located within 1 mm of the manually annotated contour, 96% within 1.5 mm, and 98% within 2 mm. The average of the maximum errors in each vertebra was 1.53 mm.

That fractures were given a higher weight when building the model makes that the performance does not decrease much when segmenting fractured vertebrae. The mean error for the fractures was 0.54 mm. 86% of the points where 1 mm within the real contour in fractures and 94% within 1.5 mm.

It is also interesting to judge the results visually, especially in cases with osteophytes and fractures, in which the contour may be oversmoothed. Some sample images are shown in Fig 5. The model has difficulties to approximate the ground truth when the latter takes abrupt turns (like around osteophytes), but otherwise works quite well, as the quantitative results suggest.

It is also important to mention that the results are not very sensitive to changes in the different tuned parameters. The increments in the mean error when changing the main parameters are:

1. for the profile length used for building the model: 4 μm in a 4-pixel-wide region around the chosen value.
2. for the profile length used in the contour search: 6 μm in a 4-pixel-wide region around the chosen value.

Table 1
Comparison of the Results from This Study and From the Literature

Authors	Modality	Error Measure	Results
Zamora et al	X-rays	Point-to-point average	≤6.4 mm (50% of cases)
Smyth et al	DXA	Point-to-line RMS	≤1.23 in 95% of healthy ≤2.24 mm in 92% of fractures
de Bruijne et al	X-rays	Point-to-line average	1.4 mm (healthy and fractures)
Roberts et al	DXA	Point-to-line average	0.70 mm in healthy, 1.23 mm in fractures
Roberts et al	X-rays	Point-to-line average	0.64 mm in healthy, 1.06 mm in fractures
Roberts et al	DXA	Point-to-line average	0.69 mm in healthy, 0.96 mm in fractures
This study	X-rays	Point-to-line average	0.47 mm in healthy, 0.54 mm in fractures

3. for the cutoff Mahalanobis distance, 25 μm in a 1-unit-wide region around the chosen value.

These values suggest that the method is robust against non-optimal choices of the parameters. The cutoff Mahalanobis distance is the parameter that affects the results the most, representing a tradeoff between freedom (better approximation to outliers) and safety (lower likelihood of implausible shapes). The profile lengths affect mostly the convergence speed.

DISCUSSION

In the current setup, higher weights were given to fractures when building the model, thus lowering the overall mean performance but increasing the accuracy in the most difficult and important cases. The weighting lowers the largest errors, too. In spite of that, an average point-to-line error lower than 0.5 mm was achieved. A total of 95% of the errors was lower than 1.36 mm. The performance is higher than those of the most recent systems in literature (Roberts et al [11] recently achieved a mean error in normal and fractured vertebrae equal to 0.64 mm and 1.06 mm, respectively, with 95% and 87% of the points 2 mm within the contour), but requires an expert to mark six points. Moreover, the error is below the intrauser variability limits for the six points described previously (6), which are slightly over 1 mm RMS (0.68 mm RMS for the method presented in this article). Even if this is not a fair comparison, as the variability in the full contour would be expected to be lower, it gives and idea of the achieved accuracy.

If the results in fractures are analyzed separately, the performance of our method does not decrease as much as those of the other systems in literature. Roberts et al (11) achieve a mean error in fractures equal to 1.06 mm, with 87% of the points 2 mm within the real contour. The mean error of our method in fractures is 0.54 mm, with 97% of the points within 2 mm. All the mentioned results are summarized in Table 1.

That the six points and the full contour were annotated in different passes implies that the solution is not constrained to pass through the six landmarks. This adds flexibility to the shape model, making it possible for it to fit a wider range of different shapes. That different radiologists made the annotations also adds some interesting variability to the model; if the segmentation of the vertebrae with this input data is successful, the resulting algorithm will be quite robust.

Even though the algorithm works well in general, osteophytes are usually detected but may be oversmoothed (Fig 3, Fig 5a). This can be clearly appreciated in Fig 6, where the mean position error is plotted against the landmark number. The curve, on which the points corresponding to the six landmarks have been highlighted, peaks clearly just after the third landmark and just before the fourth, which are the typical locations of osteophytes. The curve has local minima around the points corresponding to the landmarks, except for the middle point of the lower endplate. This is possible because the six landmarks are not constrained to be on the contour. The mean distances from these points to the contour are marked with crosses in the same figure.

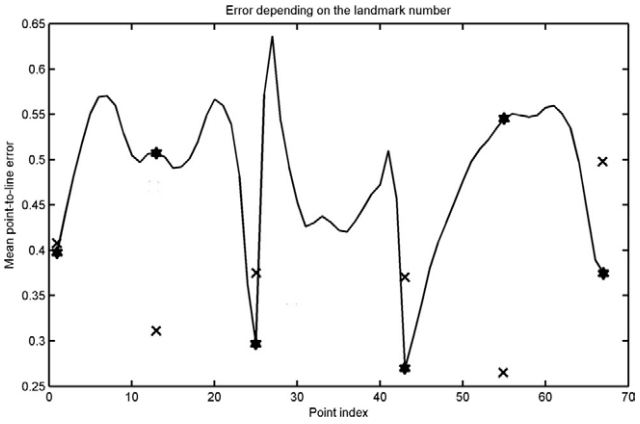


Figure 6. Error depending on the point number. The points corresponding to the six landmarks are marked with a star. The distance from the manually placed landmarks to the true contour are marked with crosses.

The easiest way of improving the segmentation around the osteophytes would be to place additional control points on their tips (as it is done in the NHANES II database of the National Institutes of Health) (16). This just represents two more points to be annotated manually for each vertebra. The only difference would be that the dimensionality of the shape vectors for the landmarks would be increased by four: two points times two coordinates. This approach was not investigated because these landmarks were not available in our database. However, using six landmarks that are part of a standard procedure (2,3) has the advantage that the method can be immediately applied to many existing datasets that have already been annotated for fracture assessment.

Another possible way to improve the segmentation around the osteophytes would be to collect more training cases in which they are present. It would also be possible to give a higher weight to such cases, which worked well with the fractures, but it would require an osteophyte-labeled training set (unfortunately unavailable). A third option could be to allow more flexibility to the ASM around the typical locations for osteophytes, for example by modifying the contour with a smooth curve to fit the points. A more complex solution could be to improve the correspondence of the landmarks around the osteophytes. Minimum description length, in which the best hypothesis for a given set of data is the one that leads to a largest compression, can be used for choosing corresponding points (17,18).

The possibility of having separate models for fractures and healthy vertebrae was also considered, but preliminary experiments showed that, even if it was known whether the vertebra is fractured, the results improved only slightly. This does not make up for the expense of having to decide whether a vertebra is fractured or not, because a mistake would lead to using the wrong model and therefore to large errors. In a similar way, preliminary experiments suggested that it was better to mix all the L1 to L4 vertebrae in a single model; the benefit of having more training cases makes up for having a customized model for each vertebra, because they are very similar in shape.

Increasing the weight of the fractured vertebrae or osteophytes during the training so that they contribute more to the model is a heuristic solution that improves the results for these cases, but it is not optimal in any sense. It would be desirable to derive values of the weights that would lead to an "optimal" result, for example minimizing a cost function that considers the importance of fractures over normal vertebrae.

CONCLUSION

A semiautomatic vertebra segmentation method based on a conditional shape model is presented in this article. Estimation of osteoporosis has been the motivation of this work, but the method can be used to quantify other pathologies, for instance disk space narrowing or scoliosis.

The method is based on a model which is built based on the position of six landmarks that a radiologist places on the image, in connection with a set of manually annotated example images. The method provides the full vertebral contour and can therefore measure more subtle shape variations. This is especially important in clinical trials, in which the algorithm could potentially reduce the required number of participants and follow-up time. It addition, the information from the whole shape may lead to earlier diagnosis and thus improved care for individual patients.

Compared with other segmentation tools described in the recent literature, our method outperforms them with the only requirement of an expert marking six points on each vertebra. The improvement is much larger when it comes to fractures, which are the most important cases, thanks to the good initialization and the constraints from the conditional model.

ACKNOWLEDGMENTS

The authors would like to say thanks to Claus Christiansen, Paola C. Pettersen, Qing He, and Jianghong Chen

for providing us with the images and manual annotations, and also to Center for Clinical and Basic Research A/S for the funding.

REFERENCES

1. Genant HK, Guglielmi G, Jergas M. Bone densitometry and osteoporosis. Springer, Berlin, 1997.
2. Black DM, Palermo L, Nevitt MC, et al. Comparison of methods for defining prevalent vertebral deformities: the Study of Osteoporotic Fractures. J Bone Miner Res 1995; 10:890–902.
3. Genant HK, Wu CY, van Kuijk C, et al. Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 1993; 8:1137–1148.
4. McCloskey EV, Spector TD, Eyres KS, et al. The assessment of vertebral deformity: a method for use in population studies and clinical trials. Osteoporos Int 1993; 3:138–147.
5. Ferrar L, Jiang G, Admas J, et al. Identification of vertebral fractures: an update. Osteoporos Int 2005; 16:717–728.
6. Smyth P, Taylor C, Adams J. Vertebral shape: automatic measurement with active shape models. Radiology 1999; 211:571–578.
7. Zamora G, Sari-Sarrafa H, Long R. *Hierarchical segmentation of vertebrae from X-ray images.* In: Sonka M, Fitzpatrick M, eds. Medical Imaging: Image Process. Proc SPIE, SPIE Press, 2003; 5302:631–642.
8. de Bruijne M, Nielsen M. *Image segmentation by shape particle filtering.* In: Kittler J, Petrou M, Nixon M, eds. ICPR, IEEE Computer Society Press, 2004 (III);722–725.
9. de Bruijne M, Lund MT, Pettersen PC, et al. *Quantitative vertebral morphometry using neighbor-conditional shape models.* In: Larsen R, Nielsen M, Sporring J, eds. MICCAI. Volume 4190 of LNCS, Springer, 2006; 1–8.
10. Roberts MG, Cootes TF, Adams JE. *Vertebral shape: automatic measurement with dynamically sequenced active appearance models.* In: Duncan J, Gerig G, eds. MICAAI. Volume 3750 of LNCS, Springer, 2005; 733–740.
11. Roberts MG, Cootes TF, Adams JE. Automatic segmentation of lumbar vertebrae on digitised radiographs using linked active appearance models. Proc Med Image Understanding Analysis 2006; 1:120–124.
12. Roberts MG,Cootes TF, Adams JE. *Improving the segmentation accuracy of fractured vertebrae with dynamically sequenced active appearance models.* In: Proceeding of MICCAI Conference Workshop on Joint and Bone Disease 2006; 1:1–8.
13. Goodall CR. Procrustes methods in the statistical analysis of shape. J R Stat Soc 1991; 53:285–339.
14. Cootes T, Taylor C, Cooper D, et al. Active shape models—their training and application. Comput Vis Image Underst 1995; 61:38–59.
15. Cootes T, Taylor C, Lanitis DH, et al. *Building and using flexible models incorporating grey-level information.* Proce Fourth Int Conf Computer Vision 1993; 242–246.
16. Long LR, Antani S, Lee DJ, et al. Biomedical information from a national collection of spine x-rays: film to content-based retrieval. Proceedings of the SPIE, 2003; 5033:70–84.
17. Davies RH, Twining CJ, Cooter TF, et al. A minimum description length approach to statistical shape modeling. IEEE Trans Med Imaging 2002; 21:525–537.
18. Thodberg HH. Minimum description length shape and appearance models. Proc Inf Process Med Imaging 2003; 18:51–62.